Methodological considerations for large-scale breath analysis studies: lessons from the U-BIOPRED severe asthma project.

Target journal: Journal of Breath Research

Waqar M. Ahmed[1], Paul Brinkman[2], Hans Weda[3], Hugo H. Knobel[4], Yun Xu[5], Tamara M. Nijsen[3], Royston Goodacre[5], Nicholas Rattray[6], Anton Vink[3], Giorgio Pennazza[7], Paulo Montuschi[8], Peter J. Sterk[2], Stephen J. Fowler[1, 9], on behalf of the U-BIOPRED study group

Affiliations: 1. School of Biological Sciences, University of Manchester, United Kingdom. 2. Department of Respiratory Medicine, Academic Medical Center (AMC), University of Amsterdam, The Netherlands. 3. Philips Research, Royal Philips B.V., The Netherlands. 4. Philips Innovation Laboratories, Philips Lighting B.V., The Netherlands. 5. School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, United Kingdom. 6. Department of Environmental Health Sciences, Yale School of Public Health, Yale University, New Haven, United States. 7. Unit of Electronics for Sensor Systems, Faculty of Engineering, Campus Bio-Medico University of Rome, Rome, Italy. 8. Department of Pharmacology, Faculty of Medicine, Catholic University of the Sacred Heart, Rome, Italy. 9. Manchester Academic Health Sciences Centre, Manchester University NHS Foundation Trust, United Kingdom.

Correspondence: stephen.fowler@manchester.ac.uk

Abstract

Methods for breath sampling and analysis require robust quality assessment to minimise the risk of false discoveries. Planning large scale multi-site breath metabolite profiling studies also requires careful consideration of systematic and random variation as a result of sampling and analysis techniques. In this study we use breath sample data from the recent U-BIOPRED cohort to evaluate and discuss some important methodological considerations such as batch variation and correction, variation between sites, storage and

transportation, and inter-instrument analytical differences. Based on this we provide a summary of recommended best practices for new large scale multi-site studies.

Keywords

Multi-site studies, breath analysis, metabolomics, quality assurance, quality control, U-BIOPRED

1. Introduction

Studies where 'omics technologies are employed have proven valuable in unravelling complex biological mechanisms and for the discovery of important markers for inflammatory lung disease[1]. However, for studies to produce clinically meaningful results, a large number of samples are typically required to provide statistical power and hence reduce risk of false discoveries[2]. Robust sampling and analysis methods are equally important, especially where additional sites and instruments will be prone to increased systematic variation.

In metabolomic studies, a standardised and well-controlled sampling and analysis approach must be adopted throughout the experimental pipeline [3–6]. The use of metabolomic workflows in breath research is challenging as breath samples typically contain high intra- and inter-sample variation, a number of sampling variables (such as diet and medical history), and the presence of exogenous artefact or contaminant metabolites. As a result, and because of limited long term storage options, it is usually not possible to form pooled quality control samples of breath to evaluate analytical variation. In addition, breath sample analysis can result in a high number of features which often exceeds the number of samples. This dominant effect, which may be specific to a single site or instrument, can influence downstream data analysis[7]. It is therefore important to consider additional quality assessment procedures in breath research where large samples sizes are used. Studies with multiple sites can introduce new variation from inconsistent sampling techniques, sample storage and transportation, and lack of control samples. Methodological considerations have previously been reported for the collection and analysis of breath samples, and international task

forces have recommended breath sampling and analysis standardisation[8–12]. However, specific guidance for large scale multi-site studies remains limited.

In this work we aim to address methodological considerations for large scale multi-site studies that may otherwise be overlooked throughout collection and analysis of breath samples, and illustrate these using data from the U-BIOPRED (Unbiased BIOmarkers in PREDiction of respiratory disease outcomes) severe asthma cohort study. Specifically, we will investigate methods to correct for batch effects, variation due to confounding factors using known asthma volatile organic compounds (VOCs), and congruence between instrument data.

2. Materials and Methods

2.1. Study information

The U-BIOPRED study is a large scale European multi-site study with the objective to reveal novel phenotypes and therapeutic targets in severe asthma[13]. Several 'omics technologies were employed in a systems biology approach using samples including blood, urine, and breath from adult and paediatric (school-aged and pre-school children) patient cohorts. Extensive clinical data and patient-reported outcomes were also collected. In this work we will focus on data from the breath volatilomics part of the study.

2.2. Breath sampling and analysis

After five minutes breathing room air through a VOC filter (A2, North Safety, Middelburg, Netherlands), participants were asked to breathe a single vital capacity into a ten litre Tedlar® bag (SKC Inc, Eighty Four, PA, United States) via a three way valve[14]. Within ten minutes, the mixed expiratory breath sample was sequentially purged onto two sorbent tubes containing Tenax GR (stainless steel, 6mm x 7", Gerstel, Mülheim an der Ruhr, Germany) by using a peristaltic pump at a flow rate of 250 ml/min. After local sampling, the tubes were transported by air or road to the Academic Medical Centre Amsterdam for central analysis and distribution.

The first sample was used for eNose analysis. VOCs were released from the sorbent tube using a thermal desorption oven (TDS 3, Gerstel, Mülheim an der Ruhr, Germany), after which the sample was transferred into a Tedlar bag with nitrogen as carrier gas. Subsequent analysis was carried out by a composite eNose platform. The eNose platform consisted of four eNoses from four different brands, using distinct sensor technologies: 1) Cyranose C320 using carbon black-polymer sensors[15], 2) Tor Vergata eNose using quartz crystal microbalances (QMB) covered with metalloporphyrins[16], 3) Common Invent eNose using metal oxide semiconductor sensors[17], and 4) Owlstone Lonestar based on field asymmetric ion mobility spectrometry[18]. The overall system included of a collective total of 190 sensors.

The second sorbent tube was sent to Philips Research laboratories in Eindhoven, in order to be analysed centrally by thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). Samples underwent automated TD (Gerstel, Mülheim an der Ruhr, Germany) using helium as the carrier gas. The sample was then purged onto a packed liner, heated to 300 °C for 3 min, and subsequently transferred to a Tenax TA cold trap (kept at -150 °C) for 2 minutes. The cold trap was then heated to 280°C at 20°C/s, after which the sample was injected into the GC column (VF1-ms, 30 m × 0.25 mm, 1 µm, 100% dimethylpolysiloxane, Varian Chrompack, Middelburg, The Netherlands) using a splitless injection method with helium as a carrier gas at a flow rate of 1.2 mL/min. The GC oven (7890 N GC, Agilent, Santa Clara, CA, USA) was set to a ramped temperature programme with the following parameters: 40 °C (5 min hold), ramp to 300 °C at 10 °C/min (5 min hold). Compounds were then transferred to a MS with a Time-of-Flight mass analyser (LECO Pegasus 4D, LECO, Sint Joseph, MI, USA) in electron ionization positive mode (70eV), with a mass spectral acquisition range between 29 to 450 Da.

2.3. Data pre-processing and treatment

Raw GC-MS files were converted to netCDF files. Both pre-processing and statistical analyses were performed in the R software environment (R Development Core Team 2017, version 3.4.2) using relevant R packages (*XCMS, BatchCorrMetabolomics,* and *Vegan*) described in detail previously [19–21]. GC-MS samples were pre-processed to create an ion fragment data matrix for further analysis and extraction of known VOCs.

Several putative asthma-related VOCs (and other commonly reported VOCs such as C5 to C15 alkanes, acetone, isoprene, toluene, and internal standards) [22–27] were targeted for further assessment, based on predicted retention indices (using cubic spline interpolation within a retention index error threshold of ± 20) and mass fragments (extracted from Pubchem and NIST online chemical databases). The identity of compounds that met these search criteria were then confirmed by pre-processing chromatograms using AMDIS (version 2.72) and subsequently NIST14 library search.

Data from e-nose instruments were mean centred (*mean* = 0, *standard deviation* = 1). MS data were normalised by the internal standard toluene-d8, and features were range scaled. Missing values (i.e. zero values) within the GC-MS data matrix were imputed using random forest proximity[28] for multivariate analyses and samples with more than 70% missing values removed.

2.3.1.Univariate and multivariate analysis

To assess the significance of variation of VOCs between groups, non-parametric Kruskal-Wallis test was used, and if statistically significant, a post-hoc test was performed consisting of a Mann–Whitney U test between two groups with Bonferroni correction. The Kruskal-Wallis test was also used to assess the influence of sample storage on sorbent tubes for TD-GC-MS analysis. To assess variation without influence of patient factors or site variation, samples were limited to the Amsterdam site, and adults with severe asthma. These samples had a maximum storage time of 39 days.

To measure and compare the effectiveness of common batch correction methods, a distance matrix calculated using Bhattacharyya distance measure was used in principal component analysis (PCA), retaining the first 3 PCs based on scree plot analysis, where the resulting inter-batch metric was the mean Bhattacharyya distance (0 equal to no batch effect). The metric was adopted from and described in more detail by Wehrens *et al.*[21].

To measure and compare dissimilarity between instruments (i.e. pairwise analysis of GC-MS *versus* e-nose, and e-nose *versus* e-nose), a Procrustes test with principal coordinates analysis (PCoA) was performed. In summary, each instrument feature matrix for comparison (with aligned sample identifications) was converted to a distance matrix based the Jaccard measure, as described in previous studies[29,30]. Principal

coordinates analysis (PCoA) was then performed on the distance matrices retaining the first 3 PCs which explained the most variance based on scree plot analysis. As it is important to correct for the occurrence of any negative eigenvalues when using the Jaccard measure (non-Euclidean) with PCoA (Euclidean)[29], we used the Lingoes correction method to convert negative eigenvalues to non-negative where a constant value double the value of negative eigenvalues, and added to all eigenvalues resulting in a corrected non-negative eigenvector[31].

Pairwise congruence between two instruments for the same patient sample was assessed using Procrustes analysis. Briefly, an input matrix is geometrically transformed (i.e. rotating and stretching/shrinking) to find the optimal superimposition on a target matrix, and such that the sum of squared distances (m^2) is minimised (where 1 = data are different, and 0 = data are the same). In order to remove bias from a single Procrustes superimposition (non-symmetric), a Procrustean test with 1000 permutations is performed to assess the significance of the superimposition, where the correlation in Procrustes rotation is derived from the permuted symmetric Procrustes rotation ($r = \sqrt{(1 - m^2)}$). Detailed explanations and considerations of the procrustean test applied to MS data have been published [32,33].

3. Results

3.1. Sample and metadata description

A total of 298 breath samples were collected in the study, of which 164 were first visit (baseline) samples and used in this work, each sample comprising breath from a single patient, thereby reducing any confounding effect from longitudinal samples. Baseline samples included adults with mild/moderate asthma (*n* = 11), and severe asthma (*n* = 42); school aged children with mild/moderate asthma (*n* = 15), and severe asthma (*n* = 37); and pre-school aged children with mild/moderate wheeze (*n* = 21), and severe wheeze (*n* = 32). Clinical characteristics and definitions of asthma severity from within the U-BIOPRED consortium are described elsewhere[34]. The remaining samples were from healthy subjects (*n* = 6). Breath samples were collected at the following sites: Amsterdam, The Netherlands (*n* = 78); Copenhagen, Denmark (*n* = 19), and London, (*n* = 47), Manchester (*n* = 12), and Southampton (*n* = 8), United Kingdom. With the exception of

Copenhagen and Southampton sites, which comprised breath samples solely from children and adults (respectively), all sites provided a mixture of age and asthma subtype groups.

3.2. Identification of VOCs associated with asthma

To assess the variation of VOCs using sample groups (i.e. by age, site, asthma severity), we first identified known breath VOCs (acetone, isoprene, and toluene), C5-15 alkanes, and putative asthma VOCs compiled from selected studies using TD-GC-MS profiling techniques. A final target list of 53 asthma VOCs was used to search the U-BIOPRED ion fragment matrix. From this search, 42 VOCs were extracted, of which 16 had identification confirmed (see table 1 and supplementary information S1). Confirmed VOCs were classed as aldehydes (n = 2), ketone (n = 1), sulphide (n = 1), furan (n = 2), and hydrocarbons (n = 10), of which seven were branched hydrocarbons.

Table 1. *A list of VOCs putatively identified as biomarkers of asthma and extracted from U-BIOPRED breath sample GC-MS data. Also shown are their average molecular weight, base peak mass fragment, the predicted retention index, the KEGG description of that VOC, and the reference from which they were putatively identified.*

VOC	*Molecular Weight (g/mol)*	*Extracted MS fragments*	*Predicted RI*	*KEGG description*	*Reference*
Carbon disulfide	76.13	44, 76	537	Non-specific	[24]
2-butanone	72.11	43, 57, 72	577	Non-specific	[26]
2-methylpentane	86.18	57, 86	584	Lipid peroxidation	[23]
2-methylfuran	82.10	39, 53	588	Carcinogen	[22]
3-methylfuran	82.10	39, 53, 82	608	Carcinogen	[22]
Benzene	78.11	51, 78	660	Xenobiotic	[25,27]
2-methylhexane	100.20	85, 100	677	Lipid peroxidation	[23]
Octane	114.23	43, 57, 114	800	Lipid peroxidation	[23]
p-xylene	106.17	91, 106	872	Non-specific	[24,25]
Cumene	120.20	105, 120	928	Carcinogen	[24]
Octanal	128.22	128	978	Lipid peroxidation	[22]
2-methyldecane	156.31	43, 57, 128, 156	1061	Lipid peroxidation	[26]
Nonanal	142.24	57	1084	Lipid peroxidation	[22]
Dodecane	170.34	170	1200	Lipid peroxidation	[27]
2,6,11-trimethyldodecane	212.42	57, 71, 212	1275	Lipid peroxidation	[26]
2,6,10-trimethyldodecane	212.42	57, 71	1332	Lipid peroxidation	[23,26]

Known analytical artefacts including polydimethylsiloxanes, phenol, and N,N-dimethylacetamide were identified. Figure 1 shows a combined mass spectrum highlighting sampling and instrument artefacts using

the mean of sample intensities. Reproducible and stable artefact peaks have been used to calibrate an instrument for qualitative analysis, as previously shown[35].

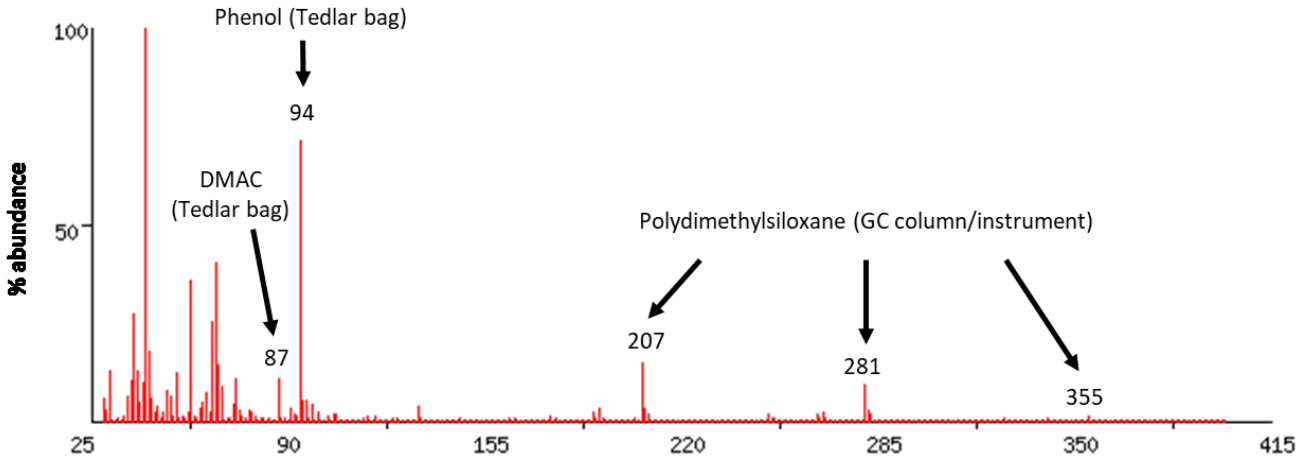

Figure 1. *A combined mass spectrum of all sample mass fragments within the MS acquisition range (29-400 Da) and their mean abundance for all samples scaled to percentage, normalised to the most abundant peak. Highlighted are Tedlar bag sampling artefacts (phenol, and DMAC), and instrument artefacts (polydimethylsiloxane).*

3.3. Patient factors and variation between sites.

Out of the previously identified VOCs, we found an increased response for dodecane (figure 2a) for the London site (adults = 4, school =16, pre-school = 20), in comparison to the Amsterdam (adults = 14, school =11, pre-school = 10) and Manchester (adults = 8, school =3, pre-school = 1) (Kruskal-Wallis $p < 0.001$). Similarly, pentane (shown in figure 2b) has shown difference (Kruskal-Wallis $p = 0.029$) in intensity between London and Copenhagen sites (Mann–Whitney U, Bonferroni corrected $p = 0.027$) within the pre-school cohort only.

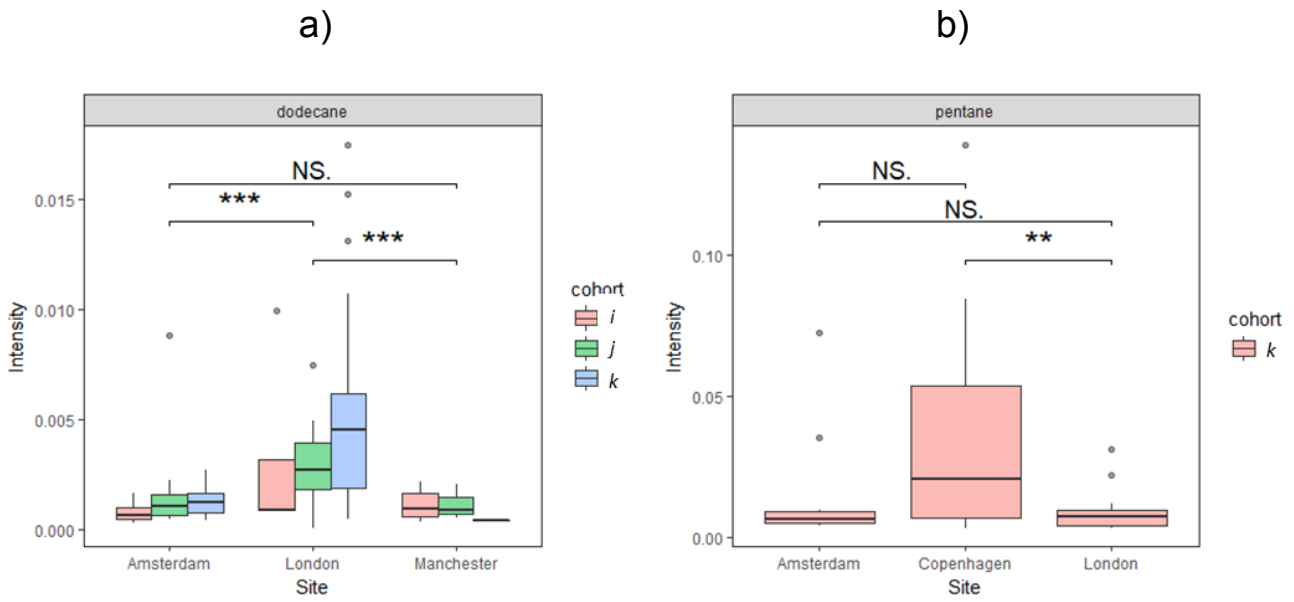

Figure 2. *Relative mass fragment intensities in patient breath samples for (a) dodecane across three sites including Amsterdam (n = 35), London (n = 40), and Manchester (n = 12), and (b) pentane across three sites including Amsterdam (n = 10), Copenhagen (n = 16), and London (n = 20). Significance between these groups are highlighted by asterisks or no significance (NS) after a Kruskal-Wallis test followed by a Mann-Whitney-U test with Bonferroni correction. Also shown for dodecane (a) are patient age groups (where i = adults, j = school children, and k = pre-school children) within each site.*

Asthma subtype was categorised as either mild/moderate or severe, as defined by IMI[34]. Using the adult cohort only, we found octanal to be increased (Mann–Whitney U p = 0.048) for patients with mild-to-moderate asthma (n = 11) when compared to patients with severe asthma (n = 22), as shown in figure 3.

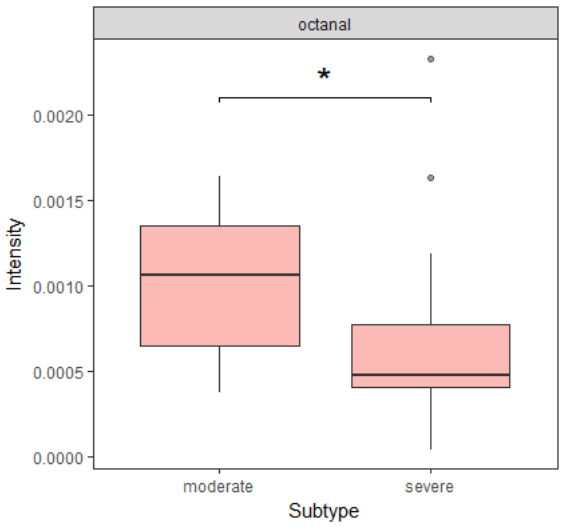

***Figure 3**. Relative mass fragment intensity showing increased octanal in patient breath samples for moderate (n = 11) compared to severe (n = 22) asthma subtypes, with a significance of p = 0.049 after a Mann-Whitney-U test.*

3.4. Variation from sample storage

After analysis of storage variation, we found no significant variation for breath samples stored for up to 39 days (Kruskal-Wallis p = 0.514). Figure 4 illustrates variation of the sample total ion count across storage duration.

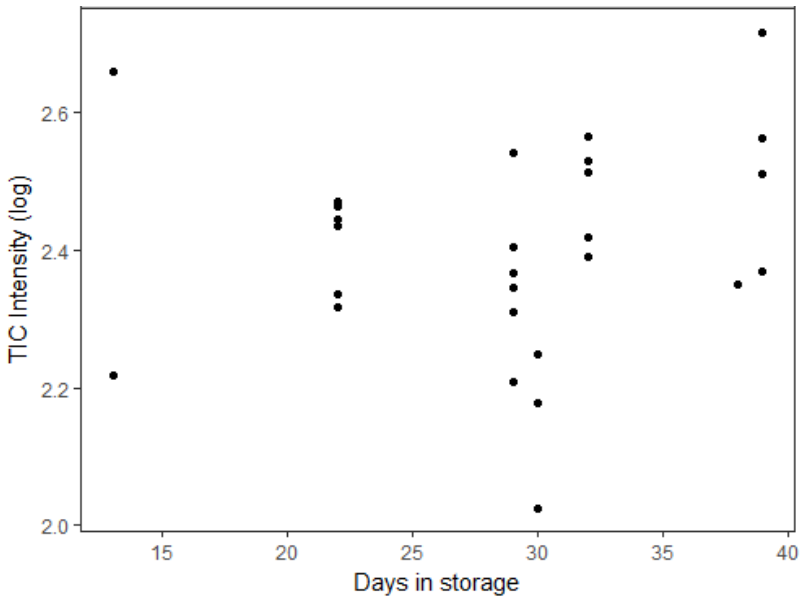

Figure 4. *Total ion count (normalised and log scaled) for samples stored for up to 39 days. Samples included severe asthma adult patients from the Amsterdam site.*

3.5. Batch correction method comparison

To compare correction methods of these batches, we adopted a method described by Wehrens *et al*.[21] In the U-BIOPRED study, breath was sampled in uncontrolled batches, where one batch may be associated to one site within a patient visit period. Samples were then analysed in controlled batches, and each sample was spiked with a four-component deuterated internal standard mixture. Mean inter-batch distances (IBD) are shown in table 2.

***Table 2**. Comparison of selected normalisation methods using the mean Bhattacharyya distance within sampling and analytical batches, where low inter-batch distance indicates larger point cloud overlap between batches, which in turn means an overall low batch effect.*

Normalisation method	Analytical batch distance	Sample batch distance	Combined average
No normalisation	0.75	0.69	0.72
Toluene-d8	0.43	0.54	0.49
Acetone-d6	2.52	2.91	2.72
Sum	0.79	0.63	0.71
Mean	0.79	0.63	0.71
Median	0.78	0.75	0.77
Sum of squares	0.87	0.90	0.89

When compared to untransformed data (combined IBD = 0.72), we show that using an internal standard (toluene-d8) improved batch correction (0.49) more than using scaling factors (between 0.71 and 0.89). In our case, scaling factors such as normalisation by sample sum or sample mean (0.71) performed better than normalisation by sample median (0.77) or sample variation (0.89). The combined IBD for acetone-d6 was 2.72. Figure 5 illustrates the sampling batch distance for data normalised by acetone-d6 (figure 5a), and toluene-d8 (figure 5b).

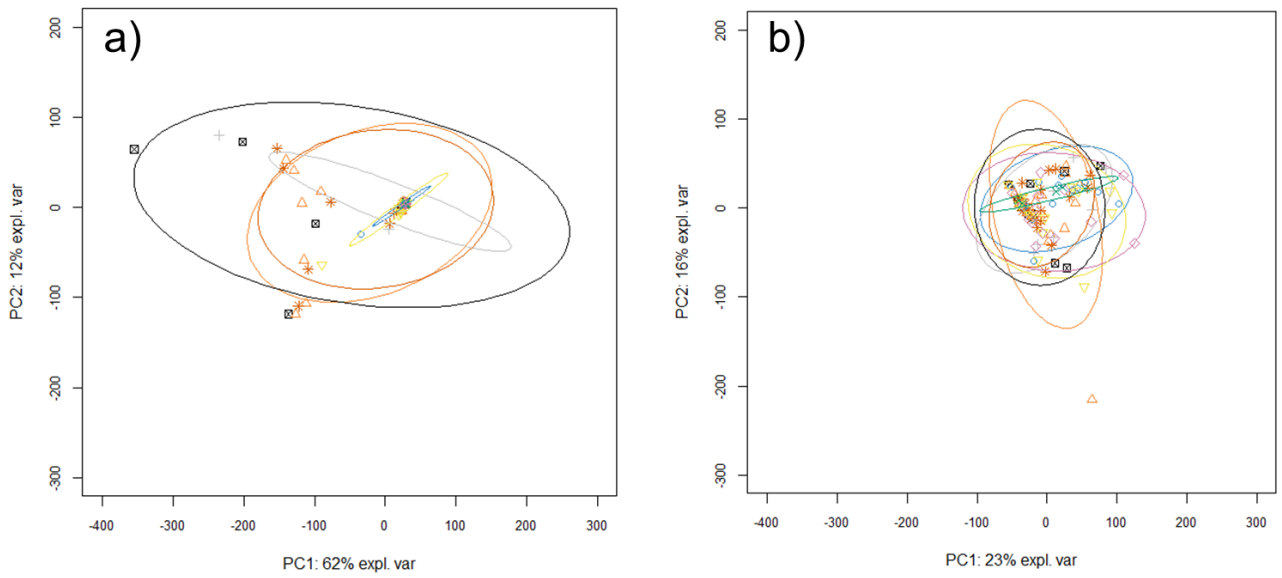

Figure 5. *Visual representations of GC-MS data normalised by internal standards a) acetone-d6, and b) toluene-d8, where batches are represented with the same colour and shape. The mean inter-batch Bhattacharyya distance is shown for each scores plot.*

3.6. Comparison between instruments

After central analysis by TD-GC-MS and an e-nose platform, Procrustes analysis was performed. Table 3 shows a pairwise matrix of correlations between instruments (Procrustes error m^2 and 95% confidence intervals shown in S2).

Table 3*. Pairwise matrix of similarity correlations derived from the Procrustes test, between GC-MS and e-nose instruments.*

	GC-ToF-MS	Lonestar	Cyranose	Tor Vergata	Comon Invent
GC-ToF-MS	–				
Lonestar	*R* 0.252 (*p* = 0.132)	–			
Cyranose	*R* 0.112 (*p* = 0.878)	*R* 0.183 (*p* = 0.204)	–		
Tor Vergata	*R* 0.211 (*p* = 0.255)	*R* 0.189 (*p* = 0.397)	***R* 0.745 (*p* = 0.001)**	–	
Comon Invent	*R* 0.173 (*p* = 0.525)	***R* 0.355 (*p* = 0.003)**	***R* 0.392 (*p* = 0.003)**	*R* 0.216 (*p* = 0.184)	–

No significant similarities were found between GC-MS and e-nose data. For e-nose comparisons, Cyranose and Tor Vergatta instruments showed correlation (r = 0.745, p = 0.001). Less strong correlations were found between Lonestar and Comon Invent (r = 0.355, p = 0.003), and between Cyranose and Comon Invent (r = 0.392, p = 0.003). To illustrate this Procrustes rotation, Figure 6 shows an example of a Procrustes superimposition plot for dissimilarity – between GC-MS and Lonestar instruments, and similarity – between Cyranose and Tor Vergatta instruments.

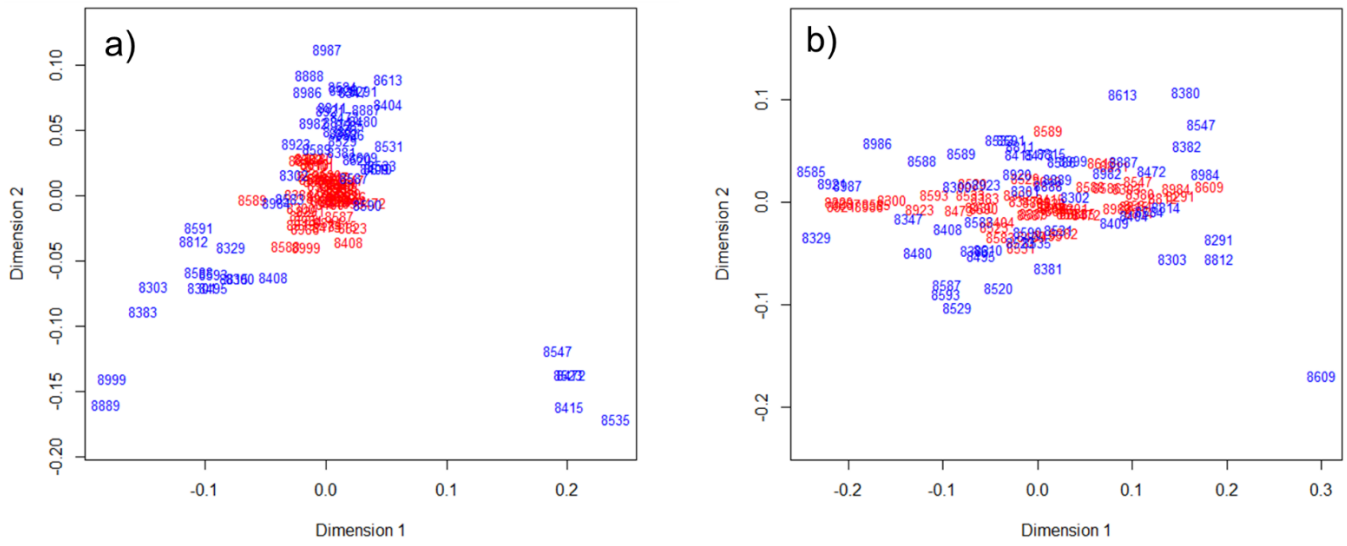

Figure 6*. Procrustes superimposition plots using distance matrices of aligned samples where a) Lonestar data (red) rotated onto GC-MS data (blue), and b) Cyranose e-nose data (red) rotated onto Tor Vergata e-nose data.*

4. Discussion

4.1. Summary of findings

In this study, we have assessed breath sample data from the recent U-BIOPRED severe asthma cohort study, and have shown possible sources of variation which must be considered when planning a large scale or multi-site study.

To achieve this, we used a targeted approach, where VOCs have been described in literature and linked to asthma, and used the KEGG database to search for their metabolic origins. It is important to stress that KEGG is curated by experts and links several chemical, biological, and pathway databases, however information of breath metabolites may be restricted as there are few databases that provide this information and feed into KEGG. With regard to VOCs found within U-BIOPRED breath data, hydrocarbons may have an exogenous source and not be linked to internal metabolism, as are many VOCs found in breath samples. A KEGG search for benzene and p-xylene revealed these VOCs as xenobiotic. Furan based cyclic compounds such as 2- and 3-methylfuran and cumene were described in KEGG as carcinogenic. These previously identified compounds may not be directly related to asthma, but may constitute epiphenomena such as differing exposures or dietary influence. Carbon disulphide, p-xylene, and 2-butanone are non-specific as they have multiple origins other than host inflammation, such as products of microbial metabolism.

Data from both KEGG and the selected studies suggest that alkanes (both saturated and unsaturated) are associated with lipid peroxidation, a hallmark of host inflammation, where volatile alkanes would be breakdown products of fatty acids (FAs). In addition, volatile aldehydes or ketones are also linked to lipid peroxidation, as carbonyl group breakdown products of FAs. Patients with severe asthma had an increased daily dose of corticosteroids, and their lower abundance of octanal when compared to mild/moderate (Figure 3) may be linked to suppression of lipid peroxidation. Although octanal may arise from other latent factors such as diet or age differences, this suppression is also evident in a previous study investigating breath VOCs from patients with COPD – which shares some clinical characteristics with severe asthma (such as airflow obstruction, frequent exacerbations, and high doses of inhaled steroids) – where a lower number of heavier VOCs were identified in comparison to non-COPD controls[36].

Similar volatile metabolic products, such as 2,6,11,15-tetramethyl-hexadecane and nonanal, have also been found in breath VOC profiles for patients with lower respiratory tract infections [37]. This may indicate the increased production of long chain unsaturated FAs due to a change in membrane fluidity, or defective efferocytosis in asthmatic patients [38]. Additionally, unsaturated FAs have been found not to induce cytokine release *in vitro* when compared to saturated FAs [39]. Shorter methylated alkanes, such as 2-methylhexane, may arise as breakdown products. It was not possible to NIST-match several methylated

alkanes because many share similar mass fragments and retention time, and therefore requires targeted analysis for accurate identification.

Sampling devices may introduce artefact VOCs depending on the type of material used, for example gas sampling bag contaminants such as phenol and N,N-dimethylacetamide [40]. Both the latter compounds were found in breath samples in this study, as breath was collected using Tedlar bags, and were removed to prevent a confounding effect during statistical analyses. Latent variables and confounders must be considered in multi-site studies, and it is important to assess any variation due to patient factors, asthma subtype, and sampling sites. Previous studies have shown variation in breath profiles due to geographical location[41,42], and these findings are supported by our analysis, where we found significant differences in levels of pentane and dodecane between Copenhagen and London sites. It is likely that the difference may be due to exogenous alkanes from the surrounding environment, but this potential contamination is especially relevant as these VOCs may also be linked to lipid peroxidation. Variation between sites can also occur from differences in sampling methods or patient demographics such as age[43], or gender[44]. In the example shown for dodecane, responses for pre-school and school aged children were increased compared to adults, and therefore age may be main contributing variable rather than a method-related effect, as the majority of school-aged patients used the same method as adult patients.

Several studies have investigated sorbent tube sample storage. We found no significant differences in our results for up to 39 days in storage, using the total ion count is a basic measurement, and we suggest using quality control samples which include known breath compounds (stored and analysed alongside breath samples) to measure variation by sample storage. Using e-nose devices to measure variation, van der Schee *et al.* found no variation where breath samples were stored for up to two weeks[45]. Kang *et al*. further investigated breath sample storage for a longer duration analysed using TD-GC-MS. They recommended storage duration may be extended up to 1.5 months[46], however this was specific to samples frozen at -80˚C and therefore not comparable to other studies where samples were refrigerated or stored at room temperature. In metabolomic experiments, -80 ˚C is recommended as metabolites do not react with each other, however this is dependent on the type of metabolites within a sample[4], and the time taken after freeze thaw and before thermal desorption.

In metabolomic studies, batch variation is intrinsic to sampling and analysis, especially for longitudinal studies. Data were batch corrected using normalisation to the internal standard. In our test we found the worst performing method was normalisation by an internal standard (acetone-d6, figure 5a) unsuitable for the sorbent material used in our analysis, as it characteristics mean it is outside the ideal Tenax GR capture range. This indicating significant "over-normalisation" of data, thereby producing a false batch effect. It is important to monitor any batch variation effect to prevent misclassification. A common batch correction method is to normalise samples by an internal standard or scaling factors[21,47]. Other methods such as ComBat or retention time alignment may also help to correct known batch variation[48].

Breath samples may be analysed centrally or on-site (i.e. near patient analysis), therefore it is useful to assess intra-sample similarity between different instruments. The expectation is instruments should show similar results for the same sample, however the instrument mechanism may overshadow any sample comparability. Compared to e-nose instruments, analysis by GC-MS produces highly resolved data based on ion fragments specific to VOCs. In contrast, e-nose data is limited to the response to several sensors and not individual VOCs. This is shown in Figure 6a, where GC-MS form additional clusters of data compared to the Lonestar instrument. Regarding e-nose comparisons, although e-nose sensing mechanisms are different, we have shown good superimposition (and therefore similarity) between Cyranose and Torr Vergata, where the reactive layer of both sensors may have similar sensing properties. We have shown that the sensing mechanism may overshadow biological differences, however further controlled experiments are required to confirm this.

4.2. Recommendations for future research

Based on our data, we provide several recommendations for breath volatile analysis for large-scale and multi-site studies from study initiation, sample collection, and sample analysis.

- To minimise variation between all sampling and analysis sites, quality assurance procedures should be incorporated into a study design. Such measures may include instrument calibration or maintenance schedule, quality monitoring, data auditing at predefined intervals with open access to metadata, and regular staff training events.

- Sampling devices (e.g. phenol and DMAC from Tedlar bags) or instrument connections (e.g. system leaks, loose connections) can be prone to contamination and therefore must be regularly serviced and/or cleaned.
- Strict storage and transportation methods should be in place where samples are stored consistently across sites without long-term storage.
- Compare data from multiple instruments to investigate similarities and differences in the analytical methods used.

Additional recommendations influenced from metabolomics-based literature useful for future studies are as follows:

- Perform regular instrument proficiency testing and use quality control samples to assess variation between instruments, as in previous studies[49,50].
- Consider data pre-treatment methods, for example, log transformation is not suitable for high RSD or missing values, and pareto-scaling can be sensitive to high sample variance[51], a common occurrence in breath analysis[7].
- Consider several models for multivariate analysis, especially for e-nose data analysis, as shown by Gromski *et al.* and Leopold *et al.* [52,53].
- Breath samples analysed with GC-MS are known to contain high multicollinearity, and any multivariate models must be optimised and validated prior to reporting[2,54,55].
- Consider using multiple databases, as highlighted by Vinaixa *et al.*[56] and the use of external standards, as recommended by the metabolomic standards initiative [57,58], to verify the identification of a VOC.

5. Conclusions

We have explored issues that may affect the results and interpretation of large-scale multi-site breath analysis studies. By using examples from the recent U-BIOPRED severe asthma study, we have shown the

importance of assessing variation which may arise between sites, patients, and instruments, with the overall aim to reduce the risk of false discoveries when interpreting results.

Acknowledgements

U-BIOPRED Consortium

We would like to thank Lieuwe Bos (Academic Medical Centre, The Netherlands), Maxim Wilkinson (University of Manchester, United Kingdom), Tom Wenseleers (University of Leuven, Belgium), and Yuanyue Wang (Royal Philips B.V, The Netherlands) for their assistance with data pre-processing.

References

[1] Wheelock C E, Goss V M, Balgoma D, Nicholas B, Brandsma J, Skipp P J, Snowden S, Burg D, D'Amico A, Horvath I, Chaiboonchoe A, Ahmed H, Ballereau S, Rossios C, Chung K F, Montuschi P, Fowler S J, Adcock I M, Postle A D, Dahlen S-E, Rowe A, Sterk P J, Auffray C and Djukanovic R 2013 Application of 'omics technologies to biomarker discovery in inflammatory lung diseases *Eur. Respir. J.* **42** 802–25

[2] Broadhurst D I and Kell D B 2007 Statistical strategies for avoiding false discoveries in metabolomics and related experiments *Metabolomics* **2** 171–96

[3] Dunn W B, Wilson I D, Nicholls A W and Broadhurst D 2012 The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans *Bioanalysis* **4** 2249–64

[4] Dunn W B, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles J D, Halsall A, Haselden J N, Nicholls A W, Wilson I D, Kell D B and Goodacre R 2011 Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6** 1060–83

[5] Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner L W, Goodacre R, Hardy N W, Taylor C, Fostel J, Kristal B, Kaddurah-Daouk R, Mendes P, van Ommen B, Lindon J C and

Sansone S-A 2007 The metabolomics standards initiative (MSI) *Metabolomics* **3** 175–8

[6] Rocca-Serra P, Salek R M, Arita M, Correa E, Dayalan S, Gonzalez-Beltran A, Ebbels T, Goodacre R, Hastings J, Haug K, Koulman A, Nikolski M, Oresic M, Sansone S-A, Schober D, Smith J, Steinbeck C, Viant M R and Neumann S 2016 Data standards can boost metabolomics research, and if there is a will, there is a way *Metabolomics* **12** 14

[7] Xu Y, Fowler S J, Bayat A and Goodacre R 2013 Chemometrics models for overcoming high between subject variability: applications in clinical metabolic profiling studies *Metabolomics* **10** 375–85

[8] Horváth I, Barnes P J, Högman M, Olin A, Amann A, Antus B, Baraldi E, Bikov A, Boots A W, Bos L D, Brinkman P, Bucca C, Carpagnano G E, Corradi M and Cristescu S 2017 A European Respiratory Society technical standard : exhaled biomarkers in lung disease *Eur Respir J* **49**

[9] Herbig J and Beauchamp J 2014 Towards standardization in the analysis of breath gas volatiles. *J. Breath Res.* **8** 37101

[10] Lourenço C and Turner C 2014 Breath Analysis in Disease Diagnosis: Methodological Considerations and Applications *Metabolites* **4** 465–98

[11] Bikov A, Paschalaki K, Logan-Sinclair R, Horváth I, Kharitonov S a, Barnes P J, Usmani O S and Paredi P 2013 Standardised exhaled breath collection for the measurement of exhaled volatile organic compounds by proton transfer reaction mass spectrometry. *BMC Pulm. Med.* **13** 43

[12] Guallar-hoyas C, Turner M a and Wilson I D 2012 A workflow for the metabolomic/metabonomic investigation of exhaled breath using thermal desorption GC–MS *Bioanalysis* **4** 2227–37

[13] Shaw D E, Sousa A R, Fowler S J, Fleming L J, Roberts G, Corfield J, Pandis I, Bansal A T, Bel E H, Auffray C, Compton C H, Bisgaard H, Bucchioni E, Caruso M, Chanez P, Dahlén B, Dahlen S-E, Dyson K, Frey U, Geiser T, Gerhardsson de Verdier M, Gibeon D, Guo Y, Hashimoto S, Hedlin G, Jeyasingham E, Hekking P-P W, Higenbottam T, Horváth I, Knox A J, Krug N, Erpenbeck V J, Larsson L X, Lazarinis N, Matthews J G, Middelveld R, Montuschi P, Musial J, Myles D, Pahus L, Sandström T, Seibold W, Singer F, Strandberg K, Vestbo J, Vissing N, von Garnier C, Adcock I M, Wagers S, Rowe A, Howarth P,

Wagener A H, Djukanovic R, Sterk P J and Chung K F 2015 Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort *Eur. Respir. J.* ERJ-00779-2015

[14] Dragonieri S, Schot R, Mertens B J A, Le Cessie S, Gauw S A, Spanevello A, Resta O, Willard N P, Vink T J, Rabe K F, Bel E H and Sterk P J 2007 An electronic nose in the discrimination of patients with asthma and controls *J. Allergy Clin. Immunol.* **120** 856–62

[15] Lewis N S 2004 Comparisons between mammalian and artificial olfaction based on arrays of carbon black-polymer composite vapor detectors *Acc. Chem. Res.* **37** 663–72

[16] Di Natale C, Paolesse R and D'Amico A 2007 Metalloporphyrins based artificial olfactory receptors *Sensors Actuators, B Chem.* **121** 238–46

[17] Bos L D J, van Walree I C, Kolk A H J, Janssen H-G, Sterk P J and Schultz M J 2013 Alterations in exhaled breath metabolite-mixtures in two rat models of lipopolysaccharide-induced lung injury *J. Appl. Physiol.* **115** 1487–95

[18] Arasaradnam R P, Mcfarlane M J, Ryan-Fisher C, Westenbrink E, Hodges P, Thomas M G, Chambers S, O'Connell N, Bailey C, Harmston C, Nwokolo C U, Bardhan K D and Covington J A 2014 Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis *PLoS One* **9**

[19] Smith C A, Want E J, O 'maille G, Abagyan R and Siuzdak G 2006 XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification *Anal. Chem.* **78** 779–87

[20] Oksanen J 2015 Multivariate analysis of ecological communities in R: vegan tutorial *R Doc.* 43

[21] Wehrens R, Hageman J A, van Eeuwijk F, Kooke R, Flood P J, Wijnker E, Keurentjes J J B, Lommen A, van Eekelen H D L M, Hall R D, Mumm R and de Vos R C H 2016 Improved batch correction in untargeted MS-based metabolomics *Metabolomics* **12**

[22] Van Vliet D, Smolinska A, Van Horck M, Jöbsis Q, Rosias P, Muris J, Dallinga J, Van Schooten F-J and Dompeling E 2017 Can exhaled volatile organic compounds predict asthma exacerbations in children?

J. Breath Res. **11**

[23] Smolinska A, Klaassen E M M, Dallinga J W, van de Kant K D G, Jobsis Q, Moonen E J C, van Schayck O C P, Dompeling E and van Schooten F J 2014 Profiling of Volatile Organic Compounds in Exhaled Breath As a Strategy to Find Early Predictive Signatures of Asthma in Children *PLoS One* **9** e95668

[24] Dallinga J W, Robroeks C M H H T, Van Berkel J J B N, Moonen E J C, Godschalk R W L, Jöbsis Q, Dompeling E, Wouters E F M and Van Schooten F J 2010 Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children *Clin. Exp. Allergy* **40** 68–76

[25] Robroeks C M, Van Berkel J J, Jöbsis Q, Van Schooten F J, Dallinga J W, Wouters E F and Dompeling E 2013 Exhaled volatile organic compounds predict exacerbations of childhood asthma in a 1-year prospective study *Eur. Respir. J.* **42** 98–106

[26] Ibrahim B, Basanta M, Cadden P, Singh D, Douce D, Woodcock A and Fowler S J 2011 Non-invasive phenotyping using exhaled volatile organic compounds in asthma *Thorax* **66** 804–9

[27] Meyer N, Dallinga J W, Nuss S J, Moonen E J C, van Berkel J J B N, Akdis C, van Schooten F J and Menz G 2014 Defining adult asthma endotypes by clinical features and patterns of volatile organic compounds in exhaled air. *Respir. Res.* **15** 136

[28] Gromski P S, Xu Y, Kotze H L, Correa E, Ellis D I, Armitage E G, Turner M L and Goodacre R 2014 Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites* **4** 433–52

[29] Xu Y, Dixon S J, Brereton R G, Soini H A, Novotny M V., Trebesius K, Bergmaier I, Oberzaucher E, Grammer K and Penn D J 2007 Comparison of human axillary odour profiles obtained by gas chromatography/mass spectrometry and skin microbial profiles obtained by denaturing gradient gel electrophoresis using multivariate pattern recognition *Metabolomics* **3** 427–37

[30] AlRabiah H, Xu Y, Rattray N J W, Vaughan A a, Gibreel T, Sayqal A, Upton M, Allwood J W and Goodacre R 2014 Multiple metabolomics of uropathogenic E. coli reveal different information content in terms of metabolic potential compared to virulence factors. *Analyst* **139** 4193–9

[31] Legendre P and Anderson M J 1999 Distance-Based Redundancy Analysis: Testing Multispecies Responses in Multifactorial Ecological Experiments *Ecol. Monogr.* **69** 1–24

[32] Peres-Neto P R and Jackson D A 2001 How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the Mantel test *Oecologia* **129** 169–78

[33] Andrade J M, Gómez-Carracedo M P, Krzanowski W and Kubista M 2004 Procrustes rotation in analytical chemistry, a tutorial *Chemom. Intell. Lab. Syst.* **72** 123–32

[34] Bel E H, Sousa A, Fleming L, Bush A, Chung K F, Versnel J, Wagener A H, Wagers S S, Sterk P J and Compton C H 2011 Diagnosis and definition of severe refractory asthma: an international consensus statement from the Innovative Medicine Initiative (IMI) *Thorax* **66** 910–7

[35] Reynolds J C, Blackburn G J, Guallar-Hoyas C, Moll V H, Bocos-Bintintan V, Kaur-Atwal G, Howdle M D, Harry E L, Brown L J, Creaser C S and Thomas C L P 2010 Detection of volatile organic compounds in breath using thermal desorption electrospray ionization-ion mobility-mass spectrometry. *Anal. Chem.* **82** 2139–44

[36] Basanta M, Ibrahim B, Dockry R, Douce D, Morris M, Singh D, Woodcock A and Fowler S J 2012 Exhaled volatile organic compounds for phenotyping chronic obstructive pulmonary disease: a cross-sectional study *Respir. Res.* **13** 72

[37] Fowler S J, Basanta-Sanchez M, Xu Y, Goodacre R and Dark P M 2015 Surveillance for lower airway pathogens in mechanically ventilated patients by metabolomic analysis of exhaled breath: a case-control study *Thorax* **70** 320–5

[38] Grabiec A M, Denny N, Doherty J A, Happonen K E, Hankinson J, Connolly E, Fife M E, Fujimori T, Fujino N, Goenka A, Holden S, Tavernier G, Shah R, Cook P C, MacDonald A S, Niven R M, Dahlbäck B, Fowler S J, Simpson A and Hussell T 2017 Diminished airway macrophage expression of the Axl receptor tyrosine kinase is associated with defective efferocytosis in asthma *J. Allergy Clin. Immunol.*

[39] Gupta S, Knight A G, Gupta S, Keller J N and Bruce-Keller A J 2012 Saturated long-chain fatty acids activate inflammatory signaling in astrocytes *J. Neurochem.* **120** 1060–71

[40] Beauchamp J, Herbig J, Gutmann R and Hansel A 2008 On the use of Tedlar® bags for breath-gas sampling and analysis *J. Breath Res.* **2** 46001

[41] Amal H, Leja M, Broza Y Y, Tisch U, Funka K, Liepniece-Karele I, Skapars R, Xu Z-Q, Liu H and Haick H 2013 Geographical variation in the exhaled volatile organic compounds. *J. Breath Res.* **7** 47102

[42] Gaida A, Holz O, Nell C, Schuchardt S, Lavae-Mokhtari B, Kruse L, Boas U, Langejuergen J, Allers M, Zimmermann S, Vogelmeier C, Koczulla A R and Hohlfeld J M 2016 A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD *J. Breath Res.* **10**

[43] Phillips M, Cataneo R N, Greenberg J, Gunawardena R, Naidu A and Rahbari-Oskoui F 2000 Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress *J. Lab. Clin. Med.* **136** 243–9

[44] Das M K, Bishwal S C, Das A, Dabral D, Varshney A, Badireddy V K and Nanda R 2014 Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS *Anal. Chem.* **86** 1229–37

[45] van der Schee M P, Fens N, Brinkman P, Bos L D, Angelo M D, Nijsen T M, Raabe R, Knobel H H, Vink T J and Sterk P J 2013 Effect of transportation and storage using sorbent tubes of exhaled breath samples on diagnostic accuracy of electronic nose analysis *J. Breath Res.* **7** 16002

[46] Kang S and Paul Thomas C L 2016 How long may a breath sample be stored for at −80 °C? A study of the stability of volatile organic compounds trapped onto a mixed Tenax:Carbograph trap adsorbent bed from exhaled breath *J. Breath Res.* **10** 26011

[47] Katajamaa M and Orešič M 2007 Data processing for mass spectrometry-based metabolomics *J. Chromatogr. A* **1158** 318–28

[48] Leek J T, Johnson W E, Parker H S, Jaffe A E and Storey J D 2012 The SVA package for removing batch effects and other unwanted variation in high-throughput experiments *Bioinformatics* **28** 882–3

[49] Nakhleh M K, Amal H, Jeries R, Broza Y Y, Aboud M, Gharra A, Ivgi H, Khatib S, Badarneh S, Har-Shai L,

Glass-Marmor L, Lejbkowicz I, Miller A, Badarny S, Winer R, Finberg J, Cohen-Kaminsky S, Perros F, Montani D, Girerd B, Garcia G, Simonneau G, Nakhoul F, Baram S, Salim R, Hakim M, Gruber M, Ronen O, Marshak T, Doweck I, Nativ O, Bahouth Z, Shi D, Zhang W, Hua Q, Pan Y, Tao L, Liu H, Karban A, Koifman E, Rainis T, Skapars R, Sivins A, Ancans G, Liepniece-Karele I, Kikuste I, Lasina I, Tolmanis I, Johnson D, Millstone S Z, Fulton J, Wells J W, Wilf L H, Humbert M, Leja M, Peled N and Haick H 2017 Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules *ACS Nano* **11** 112–5

[50] van Oort P M P, Nijsen T, Weda H, Knobel H, Dark P, Felton T, Rattray N J W, Lawal O, Ahmed W, Portsmouth C, Sterk P J, Schultz M J, Zakharkina T, Artigas A, Povoa P, Martin-Loeches I, Fowler S J and Bos L D J 2017 BreathDx – molecular analysis of exhaled breath as a diagnostic test for ventilator–associated pneumonia: protocol for a European multicentre observational study *BMC Pulm. Med.* **17** 1

[51] van den Berg R A, Hoefsloot H C J, Westerhuis J A, Smilde A K and van der Werf M J 2006 Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7** 142

[52] Gromski P S, Correa E, Vaughan A A, Wedge D C, Turner M L and Goodacre R 2014 A comparison of different chemometrics approaches for the robust classification of electronic nose data *Anal. Bioanal. Chem.* **406** 7581–90

[53] Leopold J H, Bos L D J, Sterk P J, Schultz M J, Fens N, Horvath I, Bikov A, Montuschi P, Di Natale C, Yates D H and Abu-Hanna A 2015 Comparison of classification methods in breath analysis by electronic nose. *J. Breath Res.* **9** 46002

[54] Smolinska A, Hauschild A-C, Fijten R R R, Dallinga J W, Baumbach J and van Schooten F J 2014 Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis *J. Breath Res.* **8** 27105

[55] Gromski P S, Muhamadali H, Ellis D I, Xu Y, Correa E, Turner M L and Goodacre R 2015 A tutorial

review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding *Anal. Chim. Acta* **879** 10–23

[56] Vinaixa M, Schymanski E L, Neumann S, Navarro M, Salek R M and Yanes O 2016 Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects *TrAC - Trends Anal. Chem.* **78** 23–35

[57] Goodacre R, Broadhurst D, Smilde A K, Kristal B S, Baker J D, Beger R, Bessant C, Connor S, Capuani G, Craig A, Ebbels T, Kell D B, Manetti C, Newton J, Paternostro G, Somorjai R, Sjöström M, Trygg J and Wulfert F 2007 Proposed minimum reporting standards for data analysis in metabolomics *Metabolomics* **3** 231–41

[58] Sumner L W, Amberg A, Barrett D, Beale M H, Beger R, Daykin C A, Fan T W-M, Fiehn O, Goodacre R, Griffin J L, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane A N, Lindon J C, Marriott P, Nicholls A W, Reily M D, Thaden J J and Viant M R 2007 Proposed minimum reporting standards for chemical analysis *Metabolomics* **3** 211–21

Supplementary information

S1

List of Asthma VOCs from previous studies (*n* = 53)

Extracted from U-BIOPRED feature matrix (*n* = 42)

Matched with NIST ID (n = 17)

2-butanone
2-methyldecane
2-methylfuran
3-methylfuran
2-methylhexane
2-methylpentane
2,6,10-trimethyldodecane
2,6,11-trimethyldodecane
Benzene
Carbon disulfide
Cumene
Dodecane
Nonanal
Octanal
Octane
p-xylene
Tridecane

Unmatched/incorrect NIST ID (n = 25)

1,3-Dioxolane, 2-(phenylmethyl)-
1-phenyl-1-butane
2,3,6-trimethyloctane
2,4-dimethylheptane
2,6-Dimethylundecane (C13H28)
2,6-Pyridinedicarboxaldehyde (C7H5NO2)
2-ethyl-1,3-butadiene
2-ethyl-4-methyl-1-pentanol
2-ethylhexanal
2-Furoylacetonitrile (C7H5NO2)
2-octen-1-ol
2-undecenal
3,5-Dihydroxybenzonitrile (C7H5NO2)
4-Cyclopentene-1,3-dione, 4-phenyl-
5,9-Undecadien-2-one, 6,10-dimethyl-
Allyl methyl sulfide
biphenyl
benzoic acid
butanoic acid
cis-1,2-dimethylcyclohexane
Cyclohexane
Cyclohexanol 2-butyl
heptane, 2,2,4-trimethyl-
naphthalene, 2-ethenyl-
p-Cymenene (C10H12)

Not extracted (*n* = 11)

1-Dodecanol, 3,7,11-trimethyl-
2(3H)-Benzoxazolone (C7H5NO2)
2,4-dimethylpentane
3-methylpentane
4,6,9-nonadecatriene
Benzoic acid, heptyl ester (C14H20O2)
Cedrane (C15H26)
Decahydroquinoline
Hippuric acid (C9H9NO3)
Pentadecanal
Propylene glycol, monoallyl ether, acetate (C8H14O3)

S2

3 PCs	GC-MS	Lonestar	Cyranose	Tor Vergata	Comon Invent
GC-MS	–				
Lonestar	*R* 0.2521 (m_{12} 0.9364) *p* = 0.13187 SES: 1.1544 95% CI: 0.000-0.2807	–			
Cyranose	*R* 0.1122 (m_{12} 0.9874) *p* = 0.87812 SES: -1.1682 95% CI: 0.000-0.2455	*R* 0.1831 (m_{12} 0.9581) *p* = 0.2038 SES: 0.7899 95% CI: 0.000-0.2524	–		
Tor Vergata	*R* 0.2107 (m_{12} 0.9556) *p* = 0.25475 SES: 0.6238 95% CI: 0.000-0.2657	*R* 0.1891 (m_{12} 0.9642) *p* = 0.3966 SES: 0.2162 95% CI: 0.000-0.2655	*R* 0.7445 (m_{12} 0.4457) *p* = 0.000999 SES: 10.4010 95% CI: 0.000-0.2634	–	
Comon Invent	*R* 0.173 (m_{12} 0.9701) *p* = 0.52547 SES: -0.1456 95% CI: 0.000-0.2693	*R* 0.3554 (m_{12} 0.8737) *p* = 0.002997 SES: 3.4742 95% CI: 0.000-0.2704	*R* 0.3919 (m_{12} 0.8464) *p* = 0.002997 SES: 4.2358 95% CI: 0.000-0.2540	*R* 0.2157 (m_{12} 0.9535) *p* = 0.18382 SES: 0.8761 95% CI: 0.000-0.2628	–